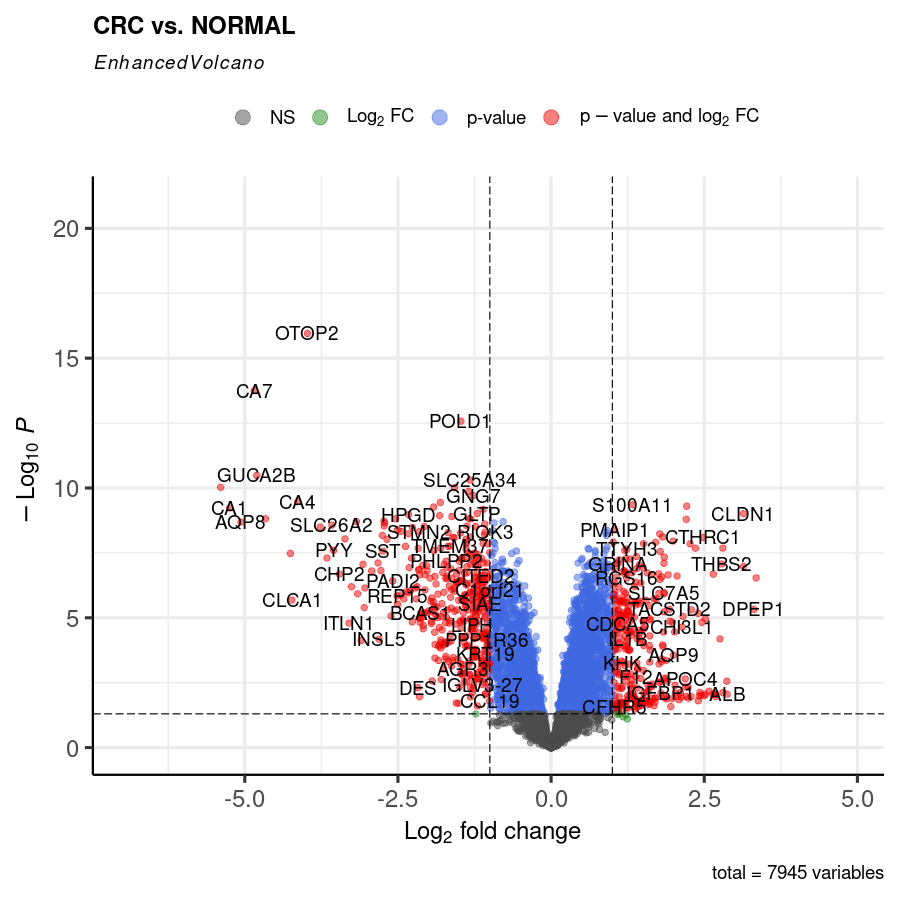
**Results:**

**Identification of differentially expressed genes**

Differential gene expression analysis revealed a total of 716 dysregulated genes in 54 CRC tumors vs normal samples (CRC tumor = 35, normal = 19). A total of 314 upregulated genes were observed with a fold change (FC) value > 1 and P-value < 0.05, while 401 downregulated genes with an FC value <-1 and P-value < 0.05 in CRC.

The dysregulated genes concerning biological and statistical significance are demonstrated in **Figure 1**. In contrast, the top 10 upregulated and downregulated genes in terms of log FC are shown in **Table 1** and **Table 2**, respectively. The top 10 upregulated genes included MMP7, DPEP1, CLDN1, SPP1, ALB, APOA2, HP, CXCL8, THBS2, and H19, and the top 10 downregulated genes included CLCA4, CA1, AQP8, CA7, GUCA2B, GUCA2A, CEACAM7, CLCA4, CA4, and OTOP2.



**Figure 1.** Enhanced volcano plot representing significant differentially expressed genes in CRC. Red dots represent upregulated and downregulated genes. Biologically significant genes are shown on the x-axis w.r.t Log2FC ± 1, and statistically significant genes are shown on the y-axis w.r.t Log10 P-value < 0.05.

**Table 1**: Differentially expressed upregulated genes

| **Genes** | **P-value** | **logFC** |
| --- | --- | --- |
| MMP7 | 2.9139149781976E-07 | 3.34801273320787 |
| DPEP1 | 4.78110613366667E-06 | 3.31039819479831 |
| CLDN1 | 9.80462599997622E-10 | 3.13884280298666 |
| SPP1 | 1.05626016710936E-07 | 3.13776252936119 |
| ALB | 0.0087945705444672 | 2.87355328101902 |
| APOA2 | 0.00278243593159611 | 2.8682673489616 |
| HP | 0.00762968662718921 | 2.80576134293485 |
| CXCL8 | 2.08544838065094E-08 | 2.80324624629453 |
| THBS2 | 8.44592655813159E-08 | 2.78446487928006 |
| H19 | 6.50145592575591E-05 | 2.75764910576636 |

**Table 2**: Differentially expressed downregulated genes

| **Genes** | **P-value** | **logFC** |
| --- | --- | --- |
| CLCA4 | 9.39943678446298E-11 | -5.39286388569494 |
| CA1 | 5.88014081870369E-10 | -5.24494580028301 |
| AQP8 | 2.06184003204157E-09 | -5.06267938261109 |
| CA7 | 1.765254609154E-14 | -4.84516708661542 |
| GUCA2B | 3.29520855046894E-11 | -4.80463212592021 |
| GUCA2A | 1.52999968250356E-09 | -4.66202473106322 |
| CEACAM7 | 3.3172660951486E-08 | -4.25681325185364 |
| CLCA1 | 2.09625636826694E-06 | -4.22885265008177 |
| CA4 | 3.38351346940158E-10 | -4.13738856316774 |
| OTOP2 | 1.11022302462516E-16 | -3.97846541040323 |

**GO term analysis of dysregulated genes**

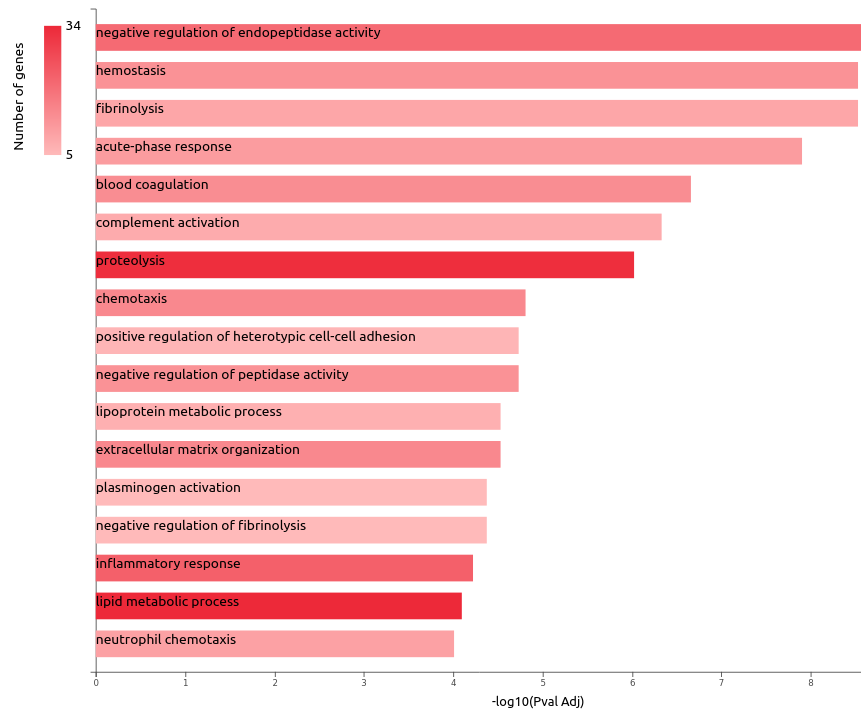
To identify GO term and KEGG pathways of dysregulated genes, GeneCodis4 was performed that identified the biological processes (BP), Cellular Component (CC), and Molecular Functions (MF) that were affected by CRC. The BP analysis demonstrated that upregulated genes were enhanced in negative regulation of endopeptidase activity, lipid metabolic process, chemotaxis, inflammatory response, neutrophil chemotaxis, proteolysis, negative regulation of peptidase activity, positive regulation of heterotypic cell-cell adhesion, negative regulation of peptidase activity, angiogenesis and lipoprotein metabolic process **(Figure 2, Supplementary Table 3)**. Nevertheless, the downregulated genes were enriched in immunoglobulin production, immune response, adaptive immune response, positive regulation of B cell activation, phagocytosis-recognition, complement activation, classical pathway, B cell receptor signaling pathway, phagocytosis engulfment, and proteolysis **(Figure 3, Supplementary Table 4)**.

The MF analysis demonstrated that upregulated genes play a role in serine-type endopeptidase activity, serine-type endopeptidase inhibitor activity, cytokine activity, CXCR chemokine receptor binding, heparin-binding, extracellular matrix structural constituent, endopeptidase inhibitor activity, growth factor activity, peptidase inhibitor activity, and serine-type peptidase activity **(Figure 4, Supplementary Table 5)**. Whilst the downregulated genes were involved in antigen binding, immunoglobulin receptor binding, extracellular matrix structural constituent, hormone activity, galactoside binding, carbonate dehydratase activity, hexosyl transferase activity, peptidase activity, guanylate cyclase activator activity, and symporter activity **(Figure 5, Supplementary Table 6)**.

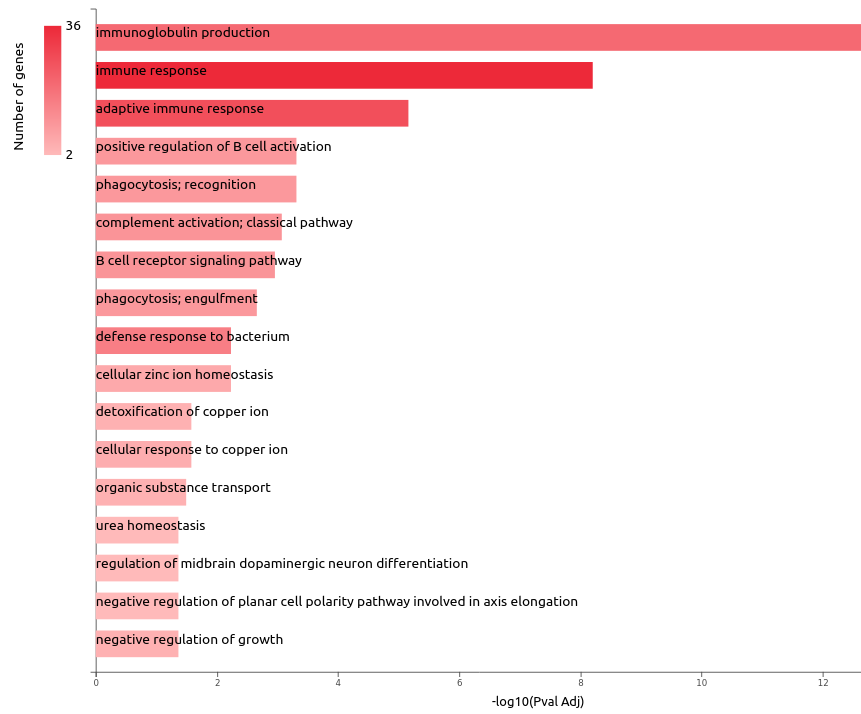
The CC analysis of upregulated genes revealed that they were majorly localized in the extracellular region, collagen-containing extracellular matrix, blood microparticle, extracellular exosome, endoplasmic reticulum lumen, extracellular matrix, high-density lipoprotein particle, platelet alpha granule, and collagen trimer **(Figure 6, Supplementary Table 7)**. In contrast, downregulated genes were mostly located in the immunoglobulin complex, collagen-containing extracellular matrix, extracellular exosome, immunoglobulin complex, circulating plasma membrane, apical plasma membrane, secretory granule, peptidase complex, microfibril, and basolateral plasma membrane **(Figure 7, Supplementary Table 8)**.

**KEGG pathway analysis of dysregulated genes**

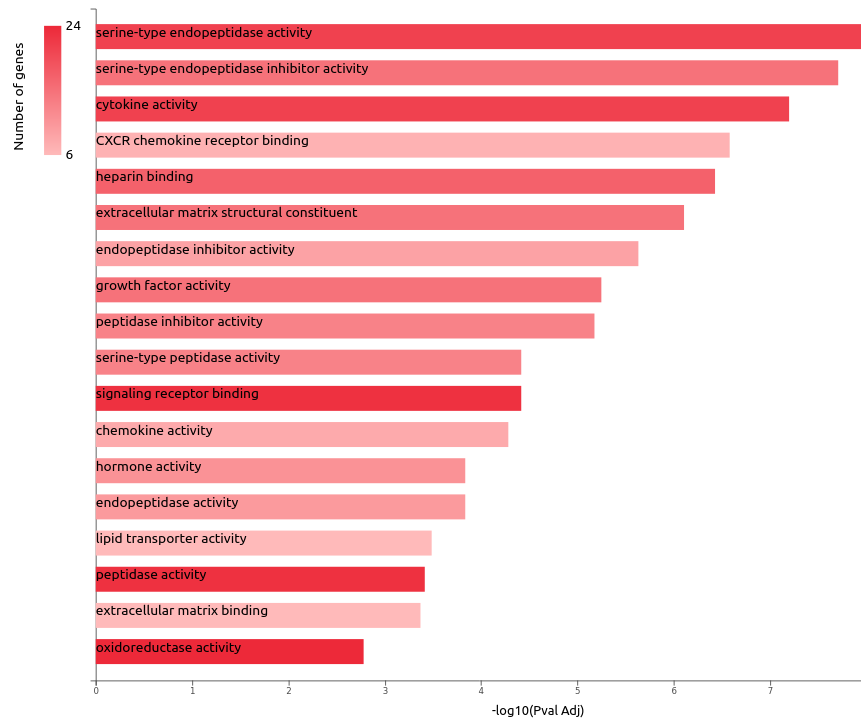
Dysregulated genes interrupted the biological pathways, which were identified through GeneCodis4. The upregulated genes were involved in overexpressed biological pathways such as complement and coagulation cascades, IL-17 signaling pathway, rheumatoid arthritis, amoebiasis, alcoholic liver disease, TNF signaling pathway, bile secretion, PPAR signaling pathway, cytokine-cytokine receptor interaction, and biosynthesis of amino acids **(Figure 8, Supplementary Table 9)**. In contrast, downregulated genes were underexpressed in bile secretion, proximal tubule bicarbonate reclamation, nitrogen metabolism, pentose and glucuronate interconversions, renin-angiotensin system, mineral absorption, protein digestion and absorption, pancreatic secretion, neuroactive ligand-receptor interaction, and aldosterone-regulated sodium reabsorption **(Figure 9, Supplementary Table 10)**.



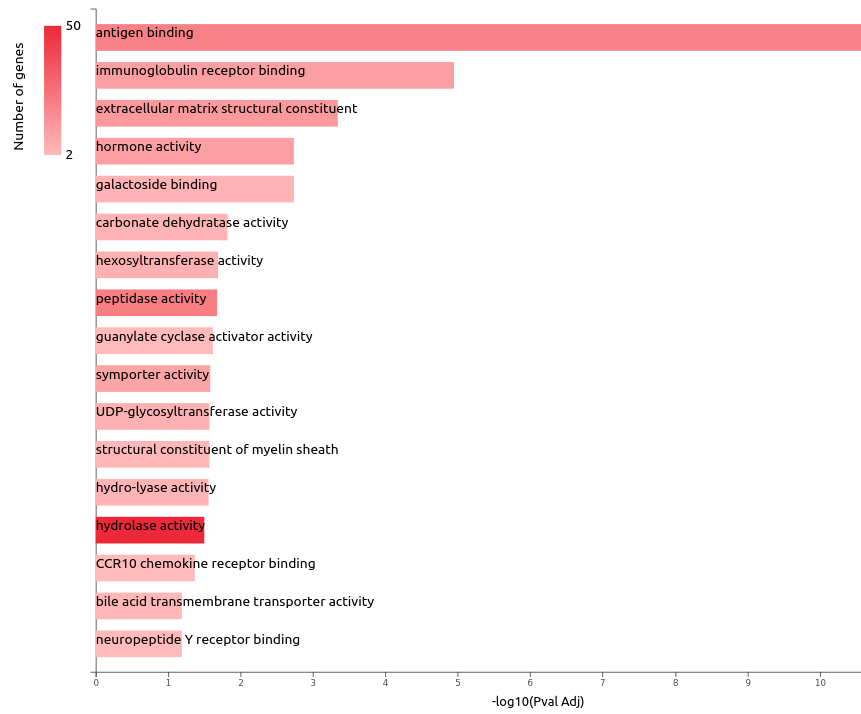
**Figure 2.** GO biological processes of upregulated genes. Barchart plot of top 20 BP in tumor vs. normal. The x-axis is the –log10(Pval Adj), while the gene number for each BP is represented on the y-axis. It represents the proteolysis and lipid metabolic process as significant BP.



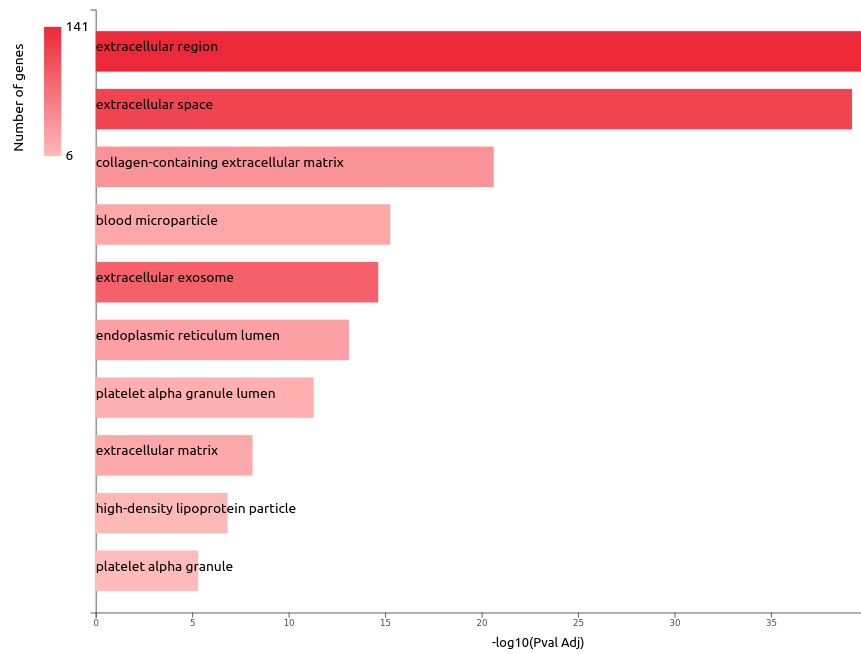
**Figure 3.** GO biological processes of downregulated genes. Barchart plot of top 20 BP in tumor vs. normal. The x-axis is the –log10(Pval Adj), while the gene number for each BP is represented on the y-axis. It represents immune response and adaptive immune response as significant BP.



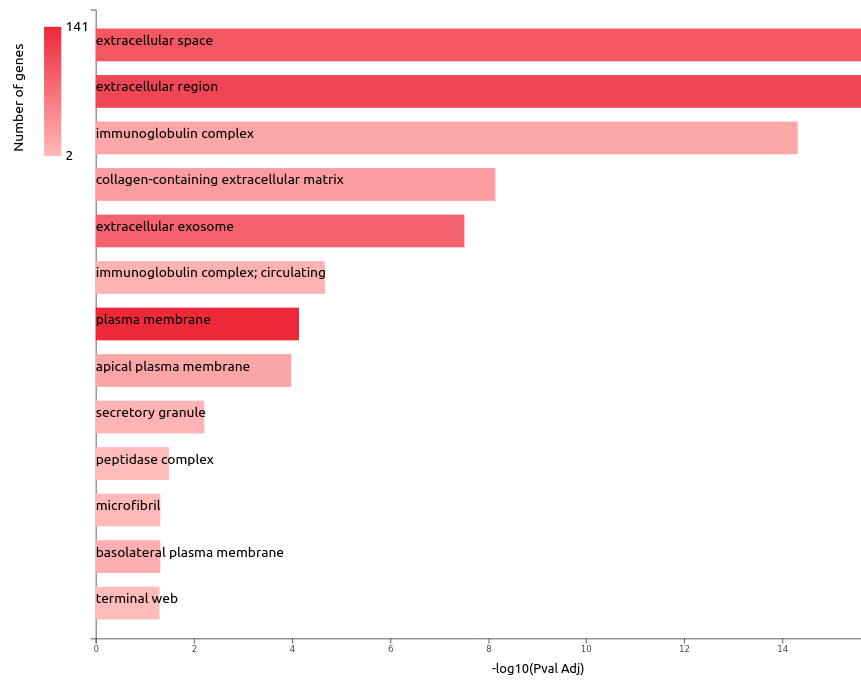
**Figure 4.** GO molecular functions of upregulated genes. Barchart plot of top 20 MF in tumor vs. normal. The x-axis is the –log10(Pval Adj), while the gene number for each MF is represented on the y-axis. It represents cytokine activity, serine-type endopeptidase activity, signaling receptor binding, peptidase activity, and oxidoreductase activity as a significant MF.



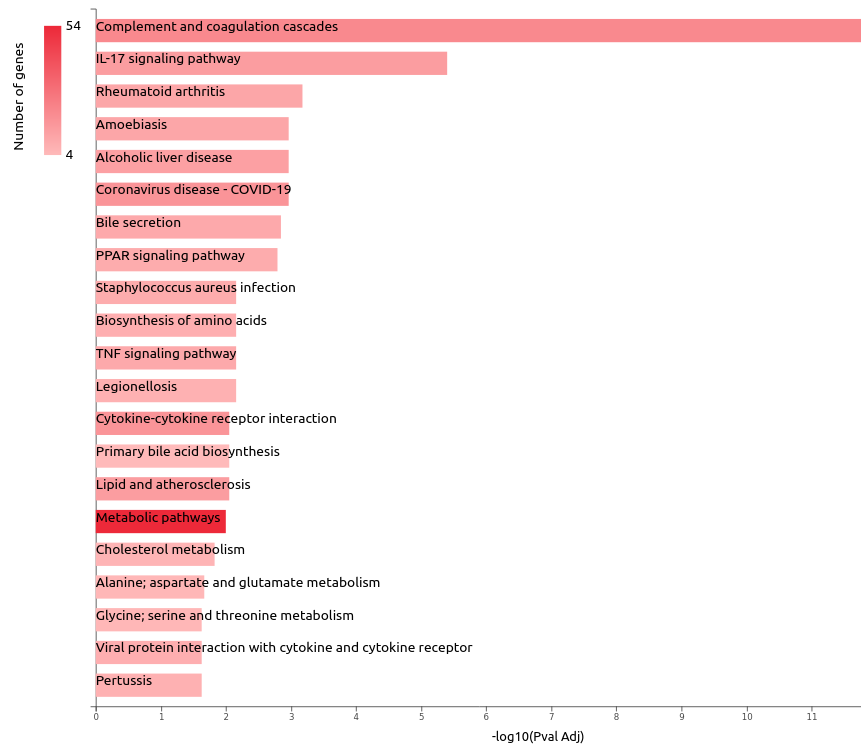
**Figure 5.** GO molecular functions of downregulated genes. Barchart plot of top 20 MF in tumor vs. normal. The x-axis is the –log10(Pval Adj), while the gene number for each MF is represented on the y-axis. It represents hydrolase activity as a significant MF.



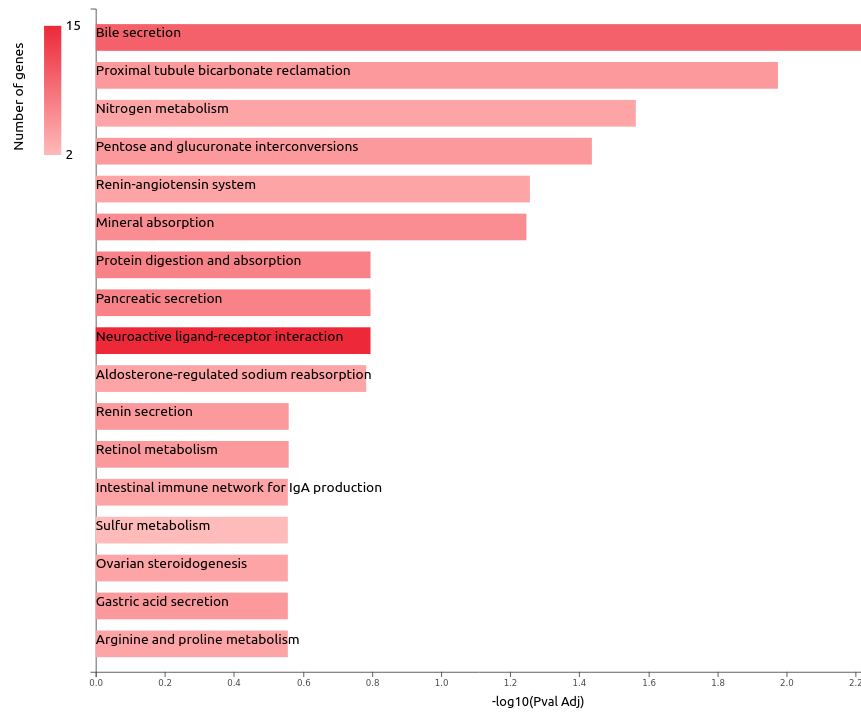
**Figure 6.** GO cellular components of upregulated genes—bar chart plot of top 20 CC in tumor vs. normal. The x-axis is the –log10(Pval Adj), while the gene number for each CC is represented on the y-axis. It represents extracellular region and space as significant CC.



**Figure 7.** GO cellular components of downregulated genes—bar chart plot of top 20 CC in tumor vs. normal. The x-axis is the –log10(Pval Adj), while the gene number for each CC is represented on the y-axis. It represents the extracellular region and plasma membrane as significant CC.



**Figure 8.** KEGG pathway analysis of upregulated genes. Barchart plot of top 20 pathways in tumor vs. normal. The x-axis is the –log10(Pval Adj), while the gene number for each pathway is represented on the y-axis.



**Figure 9.** KEGG pathway analysis of downregulated genes. Barchart plot of top 20 pathways in tumor vs normal. The x-axis is the –log10(Pval Adj), while the gene number for each pathway is represented on the y-axis.